

This article has been produced in support of and with appreciation for the efforts by Gábor Földvári of the Institute of Evolution, Centre for Ecological Research, and the Centre for Eco-Epidemiology, National Laboratory for Health Security (both located at 1121 Budapest, Konkoly-Thege Miklós út 29-33, Hungary). Through his untiring efforts, team building, and leadership, he has secured the first EU-wide team research grant. This work was supported by the National Research, Development and Innovation Office in Hungary (RRF-2.3.1-21-2022-00006) and the COST Action CA21170 "Prevention, anticipation and mitigation of tick-borne disease risk applying the DAMA protocol (PRAGMATIC), which represent the first funded efforts to apply the principles of the DAMA protocol.

An Emerging Infectious Disease Surveillance Platform for the 21st Century

Eloy Ortiz¹ and Alicia Juarrero²

1 VectorAnalytica, Miami, Florida, USA

2 VectorAnalytica, Washington, DC, USA

Corresponding author – Eloy Ortiz, VectorAnalytica, 6904 SW 88 Street, Suite F302, Miami, FL, 33156, USA,
 email eortiz@vectoranalytica.com

Abstract

Current vector surveillance programs are insufficient for coping with the emerging infectious disease crisis. In particular, current practices do not deploy sufficient information technology (IT) tools to generate actionable insights that can inform interventions and contain and mitigate the spread of vector-borne diseases. VectorAnalytica has developed a highly configurable and adaptable IT platform that imports, harmonizes, and integrates a range of data sources. The state-of-the-art backend development allows users to fully appreciate and understand the complexity of infectious disease dynamics without having to resort to additional statistical and GIS software packages. The platform yields actionable insights into outbreak patterns of vector-borne diseases that could not have been obtained without integrating pertinent contextual information at a granular spatiotemporal scale. Such integration empowers local scientists to participate actively in an ongoing manner in protecting their communities. The platform can also empower regional and national scientists and other scholars and agencies by providing them with a shared platform through which they can access continuously updated real-time data with which to formulate, validate, and update new dynamic forecasting and simulating models that support timely and appropriate decision-making and interventions on a wide range of scales. VectorAnalytica's effectiveness is described in three cases studies for which integrating and visualizing unexpected but pertinent variables (in addition to those traditionally studied for vector disease monitoring and management) yielded critical information about significant correlations between weather variables and disease incidence, and between complaints to local call centers, weather conditions, and emergent vector hotspots.

Keywords: dengue, Zika, mosquito-borne disease, vector-borne disease, rodent control, 311 hotlines, trap surveillance, data entry, mobile app, weather, census tracts, vector surveillance, infectious diseases surveillance

Introduction

Vector-borne disease outbreaks catch us unprepared time and again. Response is usually ad hoc and the status quo resumes as soon as the immediate crisis wanes. Although the problem extends most egregiously to climate change, this note turns to vector-borne diseases for use cases. As outbreaks of dengue, malaria, chikungunya, and Zika have demonstrated, these infections are exemplars of complex adaptive dynamics. Complex systems are nonliving, living, and social patterns of organization that form far from equilibrium when multiple variables and relations intertwine into complex patterns both at the individual and the collective levels in response to a variety of constraints (Juarrero, 1999, forthcoming 2023). Once formed, organisms and ecosystems, individuals and cultures adapt to and evolve in light of past experiences and current conditions. Critically, the mutually dependent relations that constitute such emergent patterns of organization and behavior display novel properties (such as herd immunity) that can be more significant than the individual events and processes that compose them. These overarching patterns and correlations cannot be clearly discerned, much less understood, and managed without taking into account pertinent contextual constraints that generate and govern them. In order to reveal the emergent properties and behavior of the complex dynamics enacted by vector-borne diseases, the framework of the IT Surveillance Platform developed at VectorAnalytica (and deployed in the use cases that follow) is informed by complexity theory.

Infectious diseases can be better contained and mitigated if they are anticipated (Brooks et al., 2014; Brooks et al., 2019; Hoberg et al., 2022). To do so, however, requires a proper monitoring and tracking system. Historically, this meant expensive state-of-the-art computers managed by experts in statistics and software. Alas, despite the explosive expansion in computational power and reduction in processing costs, IT monitoring of vector-borne diseases even in the developed world still has not lived up to its potential. Public health agencies, especially those at the municipal and county levels, inevitably rely on outdated hardware and software, and data management is highly siloed in labs, clinics, and political departments. As a case in point, systematic and digitalized data collection is rare. Integrating field data obtained by vector control departments with laboratory, clinical, and public health data is even rarer.

Meager resources of public health departments mean that locally gathered data is returned after processing with a lag of several weeks. Despite mosquito lifespans of approximately 2 weeks, laboratory analyses of larvae and blood samples are commonly weeks old by the time

the results arrive back at the local vector control or public health departments. Consequently, as was vividly exposed by the Zika and COVID outbreaks, local public health officials depend on post hoc analyses of incidence and prevalence trends. The problem with this lag time is compounded by the fact that the information provided by statewide or national organizations is aggregated and coarse grained; granular correlation between the prevalence of mosquitoes, weather conditions, and case incidence in a particular municipality and local cases with which to support decision-making is unavailable. This point can be generalized to all complex systems. Because they are exquisitely context dependent, aggregate analysis misses subtle and significant nuances of local and idiosyncratic conditions. Maps and other statistical reports generated from that information are therefore not "fit for purpose," if that phrase is understood as supporting actionable, effectual recommendations that local agencies can rely on to implement interventions and contain and mitigate an infection's spread.

To illustrate the process, we begin with data collection. Even in major US cities, data about cases and vector specimens commonly consists of a handwritten entry on a paper form. Handwriting is error fraught: answers are incomplete, illegible, misidentified, and so forth. Furthermore, paper data entry is not only inaccurate and incomplete but also has the following drawbacks:

- Paper entry usually does not include precise and automated geolocation or time stamp of the data site. As mentioned, vector control and lab data end up being aggregated, submitted, and processed with such significant time delay as to be useless for establishing medium and long-term trends, much less acting preemptively on them.
- In those rare cases in which data are entered digitally, software packages used by vector control departments, local laboratories, and various hospitals and physicians are not compatible. Comprehensive insight into current conditions becomes impossible.
- Additional relevant information about weather and climate data, population density, vegetation indices, and neighborhood socioeconomic demographics, all of which are readily available from third-party providers such as satellite feeds and census tracts, are not taken into account.
- Existing data in a range of formats (from APIs through Excel files to satellite feeds) are not incorporated, much less harmonized systematically with vector control, laboratory, and clinical data. For example,

troves of archived data in Excel format languish in desk drawers for lack of integration and processing with current data collection efforts.

- Potentially valuable geolocated and time-stamped citizen scientist reports are typically not even on the radar. If they do exist, they are often siloed in the records of the organization that sponsored the project.
- Seemingly tangential but potentially valuable and public information available from other databases such as those compiled by nationwide 311 call centers or United Way offices are not integrated for processing with vector control and clinical and lab data. We include visualizations from case studies using 311 data and census tracts in the following illustrations.

The analytic stage of existing surveillance systems is riddled with analogous deficiencies. This stage commonly involves the following steps:

- Some local communities send paper data records collected and entered manually to the nearest university or to a separate office to be "cleaned up" and then entered manually on a computer. Because of paper data entry, digitally mapping the data must then also be performed manually. Even this data, however, is rarely available at a sufficient granular resolution (read municipal, much less zip code) to support significant and effectual interventions. Even though zip code scale analysis is easily obtainable, local decision makers commonly have no access to real-time simulations and forecasts of local risk factors for that granularity.
- Academic studies that rely on data supplied by public health agencies are published with a one-year time lag and, moreover, are rarely fed back to decision makers in support of actual interventions. Transference too is lacking.
- Relevant processing requires purchasing additional software packages for statistics and mapping (Matlab, SPSS, SAS, ArcGis, Tableau, etc.). But this typically means local health agencies must hire additional personnel with the advanced math and stats skills to deploy, integrate, process, and analyze data collected for analysis by these packages. As always, pitifully meager local budgets for public health surveillance and the lack of skilled personnel in rural communities compound the problems.
- Failure to build real-time data collection and analytics into existing surveillance systems exacerbates

local bottlenecks and supply chain problems. Consequence: no state-of-the-art actionable project management recommendations based on machine learning or even simple simulations is available to local decision makers. These failures range from anticipating and monitoring inventory stock of vaccines, bed nets and repellents, fumigation chemicals, and more. As significantly, there is no way to track the effectiveness of interventions (such as comparing one form of intervention with another, much less understanding any synergies that might arise from a combination of different types of intervention).

- Current surveillance analytics are often limited to passive point maps of dated reports; these cannot be incorporated into digital processing to inform policy and decision making at the local level, much less to anticipate outbreak trajectory and severity and inform interventions.

The Bottom Line

There are structural and systematic reasons for the common and significant delays between suspected clinical or entomological events on the one hand and interventions such as vaccination campaigns or fumigation schedules on the other. This lag time in anticipation delays and obstructs containment to prevent transmission. Negative public health outcomes are real consequences of the long intervals between handwritten, paper-based data capture, processing, and returning of analytic output to local decision makers. Lag times mean unavoidable delays and consequently ineffectual interventions.

Failure to implement an accurate and efficient system with all these functionalities inevitably results in increased morbidity and mortality. Because complex systems are exquisitely context and path dependent, and because their dynamics have properties and consequences that the individual components that make them up do not, the effects of these deficiencies are often worse than anticipated.

Granular and timely insight, in real time, about current local conditions and their likely trajectory must be communicated to decision makers who can undertake appropriate and timely interventions. But real-time alert "push notifications" to the public are currently unavailable.

The consequences of communication failures can be worse than no information at all. Stale and out-of-date information can be worse than no communication at all. It misleads the community about the reality of the situation, leaving the playing field open to misinformation spread by unreliable and inaccurate sources. As noted, it makes accurate local risk assessment and actions impossible for

city and county public health managers. In consequence, medium to long-term “planning” by local authorities and the public alike becomes a series of conjectures inspired at best by historical and inaccurate anecdotes rather than data-based analysis.

Understanding the challenges to transitioning to fully digital monitoring and tracking systems is not difficult, nor need it be time consuming if set up properly from the beginning. It requires:

- Enhancing (not replacing) existing governmental and corporate tools
- Converting existing historical paper-based data into digital format
- Transitioning new data entry to digital format
- Integrating data from multiple sources into one comprehensive system while protecting privacy and ensuring data security at every scale
- Developing the IT backend that incorporates state-of-the-art mapping, statistical processing, simulation, and forecast modeling with machine-learning algorithms to eliminate the need for additional software or additional skilled personnel
- Generating forecasts and risk assessments down to the neighborhood block scale
- Visualizing analytic output of all integrated data with one click
- Empowering local authorities by placing full control of the software in their hands
- Empowering academic research to contribute actively to real-time decision-making support and interventions

The goal, of course, is to establish a shared information framework through which health-care agencies, their suppliers, and the public can more effectively monitor and track, anticipate, manage, and contain outbreaks of infectious diseases in general and of vector-borne diseases in particular. The platform must be useful at scales ranging from local public health decision makers to national organizations; it must also provide input to county and state research and support organizations and feedback from them back to inform the models and simulations. Because, to repeat, complex systems are context dependent, granular information about local conditions is necessary. Data aggregated at the national data level cannot provide insight and understanding (much less recommendations for action) about locally unique conditions.

Solutions

VectorAnalytica’s management platform for vector-borne diseases was developed to support decision making by health care agencies, their suppliers, and the communities they serve. On that platform, users

- enter official primary (vector, clinical, lab, etc.) data through mobile devices on mobile app query forms that users create on easy drag-and-drop interfaces and which they build to their specifications. To dramatically improve the quality, timeliness, and thoroughness of data collection, accurate data entry must be possible even without internet connectivity.
- integrate and analyze archived data such as third-party demographic information, satellite feeds of vegetation and weather variables, and so forth while respecting the owner’s privacy and proprietary requirements. Because VectorAnalytica software is meant to complement and not replace existing governmental or corporate surveillance efforts, its IT platform does not store third-party data.
- automate data entry of APIs as well as other tangential databases such as those of 311 or other nationwide crisis response networks.
- empower citizen scientist projects in order to generate community buy-in for improved public health surveillance.
- process and map all data without additional software packages or GIS skills. Users must also be able to disaggregate analytic output by date range, variable, correlation, and especially by local geographic area.
- produce (in real time) visualizations of statistical, mapping, and machine-learning analytics at scales ranging from neighborhood block to global levels (with the capability of disaggregation, as just mentioned, as well). Visualization significantly increases situational assessment and insight into local, current, and anticipated conditions.
- generate scientifically validated risk assessments and recommendations for local communities in real time.
- quickly disseminate information and recommendations to other public health agencies and the community at large through SMS push notifications or email.
- establish a trusted and transparent interactive communications network between authorities and the public, and between health care agencies and their suppliers.

Three Case Studies

The following are three examples of insights revealed by the VectorAnalytica platform's integrative and analytic capabilities.

Case Study 1. Zika outbreak in Miami-Dade County, 2016

At the start of the Zika virus outbreak in 2016, the vector control department of Miami-Dade County, Florida, had fewer than 30 mosquito traps placed around its 2,400 square miles. Given the impossibility of establishing

credible prior probabilities from data from such few traps in such a large geographical expanse, VectorAnalytica turned to the existing rich trove of data of "mosquito nuisance complaints" recorded on the county's 311 hotline (see Figure 1a).

Because of the low number of permanent residents in certain well-trafficked areas like Miami International Airport, however, relying on absolute numbers of 311 calls alone would not have provided actionable insight. To remedy the inability to formulate traditional indices such as the Breteau index, household indices, and others, due to the paucity of trap data, VectorAnalytica's chief scientist created

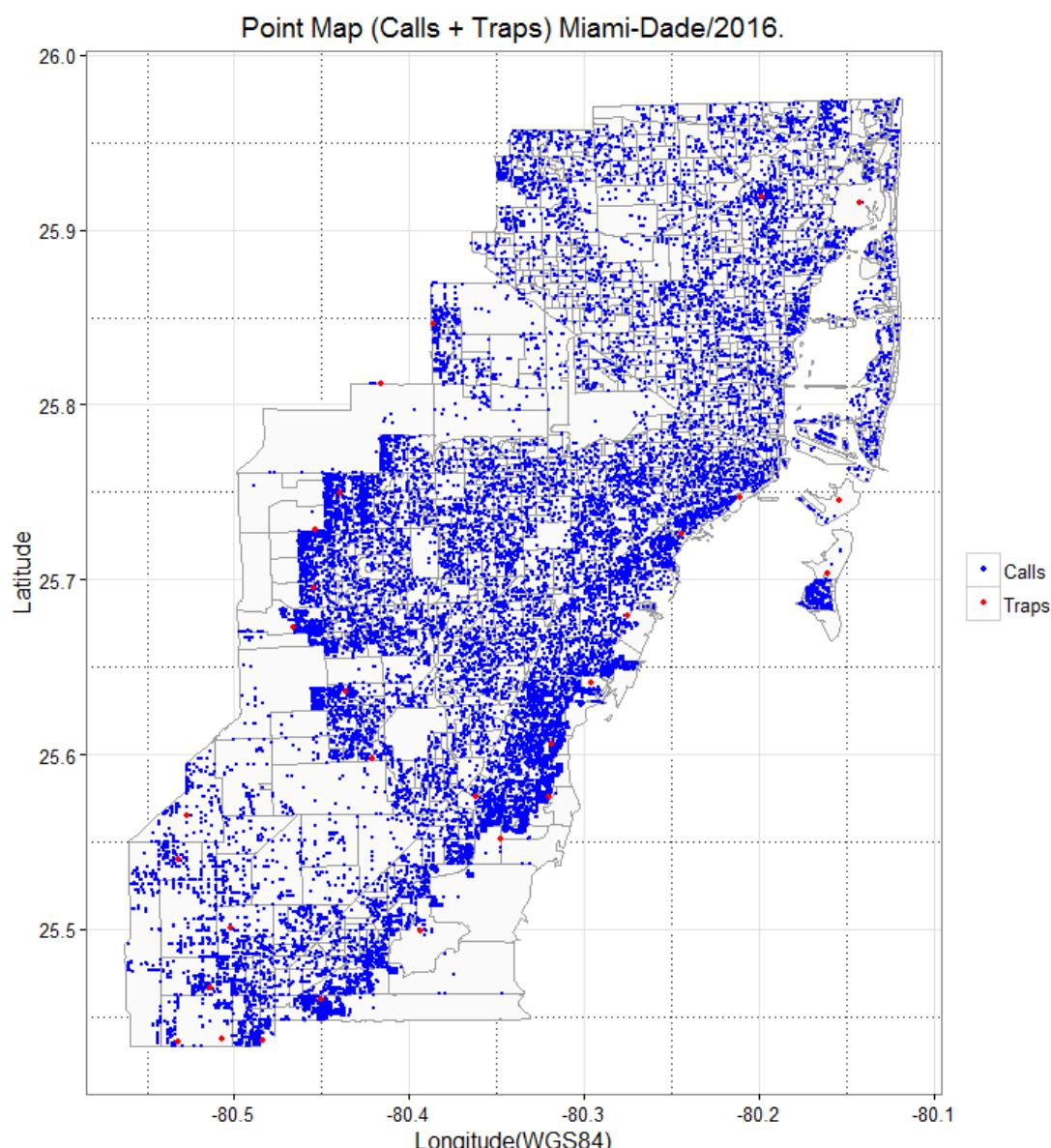


Figure 1a. Point map of Miami-Dade County, Florida, 2016. Blue dots: "Mosquitoes nuisance" complaints to the 311 hotline. Red dots: Mosquito traps.

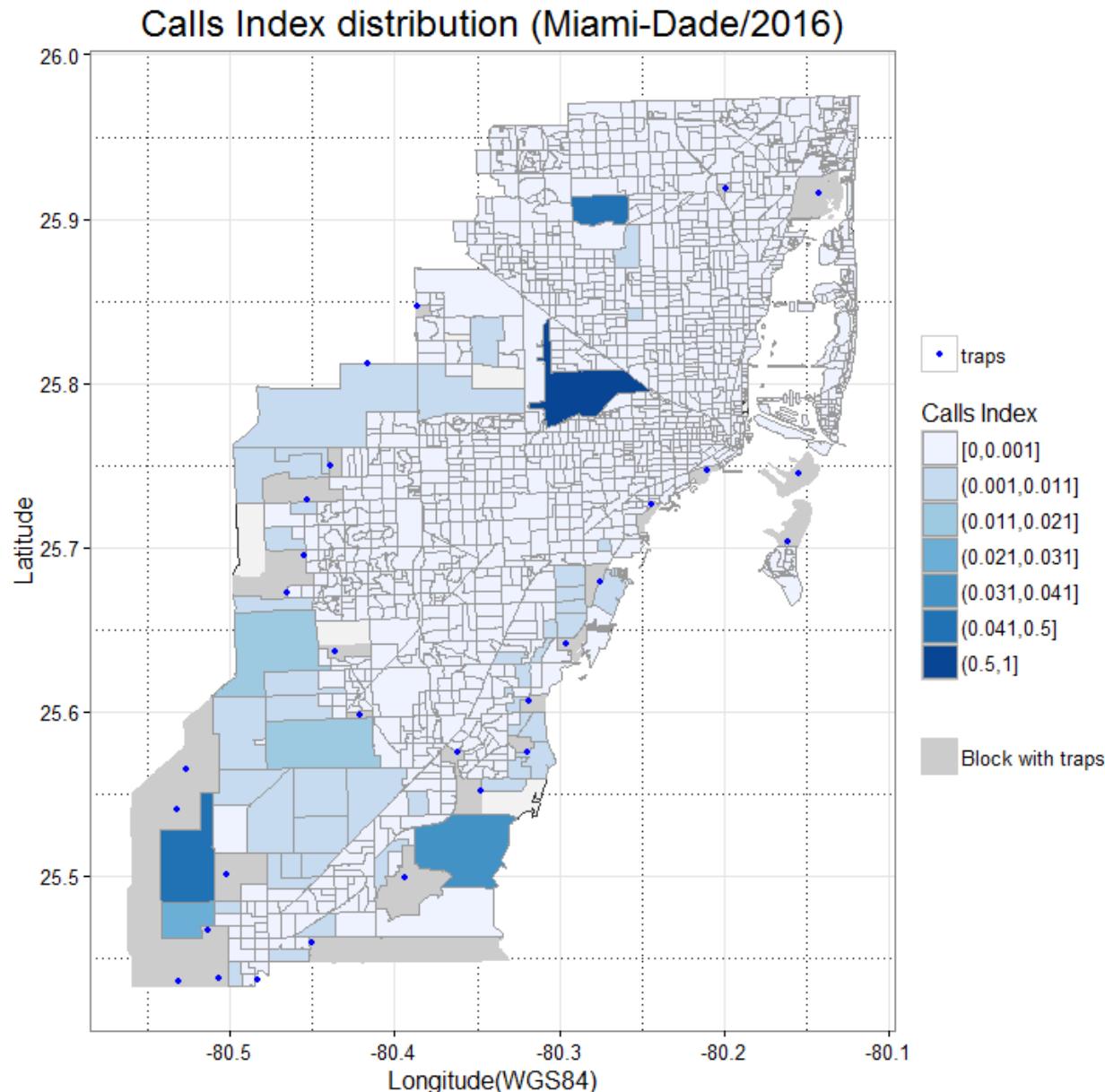


Figure 1b. Choropleth or thematic map. Call index in various shades of blue. Trap locations in gray. Miami-Dade County, Florida, 2016.

an indirect transmission risk index by integrating variables from multiple sources. The fact that Zika transmission had already begun in the county supported the assumption that a certain percentage of those phone complaints reported the presence of *Aedes aegypti* mosquitoes. To obtain that index, VectorAnalytica used a quotient obtained by dividing the number of 311 calls per block by the population density of that block (per 2010 census tract). The resulting thematic or choropleth map (see Figure 1b) uncovered differences in transmission risk in different areas of the county.

Integrating census information of residential density with location and frequency of 311 "mosquito nuisance" complaints was revelatory. As the area of the map colored in darkest blue shows, an area adjacent to the airport unmistakably emerged as the county's zone with highest transmission risk. This was initially puzzling, for it is not a densely populated residential area. However, it is the location of the Miami Intermodal Center, where the airport's major parking garages and rental car center are housed, along with the region's Amtrak and Tri-Rail hubs linking the

Ft. Lauderdale airport with the Miami airport. Major county metrorail, metrobus, and automated people-mover stations are also located there. In addition, the area has a number of small hotels and motels that host overnighting airline crew members. To complete the picture, airport authorities maintain extensive and attractive landscaping of tropical foliage, nestled into several large containment ponds. It is also an area prone to flooding and standing water after summer downpours.

The VectorAnalytica index thus revealed a realistic and significant central convergence point with respect to mosquito-borne transmission risk. Although it has few permanent residents (therefore a low denominator for the index's quotient), the area identified by the thematic map experiences a large daily volume of international and national travelers arriving at the Miami airport or transferring to or from the Ft. Lauderdale airport, from which they disperse by bus, people movers, and trains to other areas of the county, or travel beyond South Florida. A high quotient for transmission risk for these few blocks was almost obvious in retrospect.

Subsequent iterations of these maps revealed the following interesting property indirectly suggestive of the direction in which transmission would spread: Wynwood, the site of the first major Zika outbreak in the county is a nearby neighborhood accessed on the Airport Expressway (Route 112) adjacent to the Miami Rental Car Center. Census information states that the neighborhood is historically identified as Miami's Little Puerto Rico, with many longtime residents continuing to travel often to the island (which by early 2016 was already a serious focus of Zika infections). Add to that information the fact that, although still composed for the most part of rundown and abandoned warehouses and older residences, Wynwood had experienced trendy restaurants, designer shops, artists' studios, and nightspots openings in the last decade. These attractions had been drawing increasing customers from Miami Beach, a short causeway drive away. It had also become a place where travelers to and from Miami International Airport would stop, shop, and grab a bite to eat on their way to or from the beaches.

It is therefore reasonable to conjecture that those traveling from the Miami Intermodal Center to Miami Beach (or vice versa) might have become exposed to the virus during their Wynwood stop and shop. This conjecture is reinforced by a corollary, that it is likely that the Airport Expressway served as a transmission channel for contagion and spread not only to Wynwood but also to North Miami Beach, which became the second Zika hotspot in the county. It is curious to note as an aside that Route 112's terminus on the barrier island is located in an area some call Little Brazil.

Despite the absence of other more direct and pertinent data (such as species of mosquito specimens collected from numerous traps and confirmed in a lab as infected, or as gravid females—only female mosquitoes bite, etc.), and the fact that the index utilized is not a validated index, under such conditions of uncertainty, integrating and visualizing disparate sets of data (GPS-identified calls of mosquito biting combined with census information) revealed insightful information that narrowed the search for a common source of dramatically increased transmission risk to a very circumscribed neighborhood. The thematic map generated from the call index also suggested a possible channel of transmission.

Having this information earlier would have supported a far more targeted fumigation campaign (and probably as effective but less ecologically harmful and certainly less expensive) than the blanket aerial spraying the county chose to embark on instead.

Case Study 2. Mosquito-borne diseases in Costa Rica

It is common for papers reporting the results of studies of historical outbreaks of dengue and other mosquito-borne diseases to note the correlation between prevalence of clinical cases and ambient temperature (typically increasing in summer months, e.g.). When vector-control data of mosquito-borne infections is entered by hand, however, this correlation appears to be anecdotal at best.

The ability to import from OpenWeather APIs records of 23 local weather measurements in Costa Rica (temperature, humidity, dewpoint, atmospheric pressure, etc.) allowed VectorAnalytica's integrative IT platform to combine official Costa Rica public health department reports of clinical cases of mosquito-borne diseases with weather data for that country during 2016–17.

Time series visualizations of these multiple variables clearly indicate that, contrary to traditional reports in the literature, the one factor with which clinical case numbers is not systematically and universally correlated in Costa Rica is ambient temperature, reported in Sri Lanka by Goto et al. (2013). Also in contradiction to the literature (Polwiang, 2020), which rarely mentions dewpoint in connection with these infections, the correlation from these analytics does show a systematic—if time-lagged—correlation for Costa Rica between humidity and dewpoint on the one hand and clinical case increase on the other.

Comparing Zika and dengue with weather conditions uncovered that the different pathogens appear to be correlated differently with different weather conditions in different locales. This suggests a complex spatio-temporal context dependence indexed to a particular virus in a particular

microclimate and/or geographic location. It is likely, therefore, that the coevolutionary dynamics of vectors, pathogens, and hosts intertwine in a radically complex manner with local weather trends and conditions, *sui generis* vegetation cover, and so forth. This dynamic makes local vector surveillance and monitoring the key to containing and mitigating spread.

The following graphs illustrate the value of integrating disparate variables as proposed by complexity theory. As noted, the literature reports a correlation between transmission of vector-borne diseases such as dengue and Zika and annual temperature peaks. Although studies pertaining to the role of dewpoint in increases and decreases of clinical cases are rare and recent (Nygren et al., 2014), these

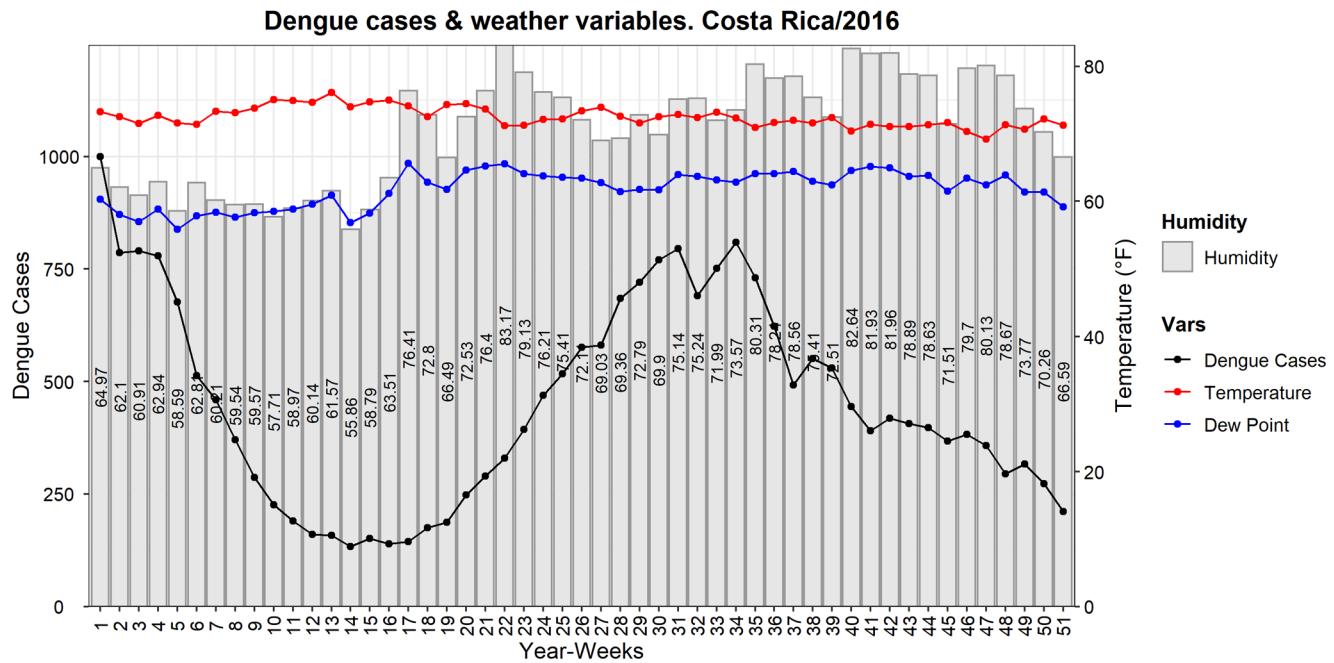


Figure 2a. Time series of dengue and weather variables for Costa Rica, 2016.

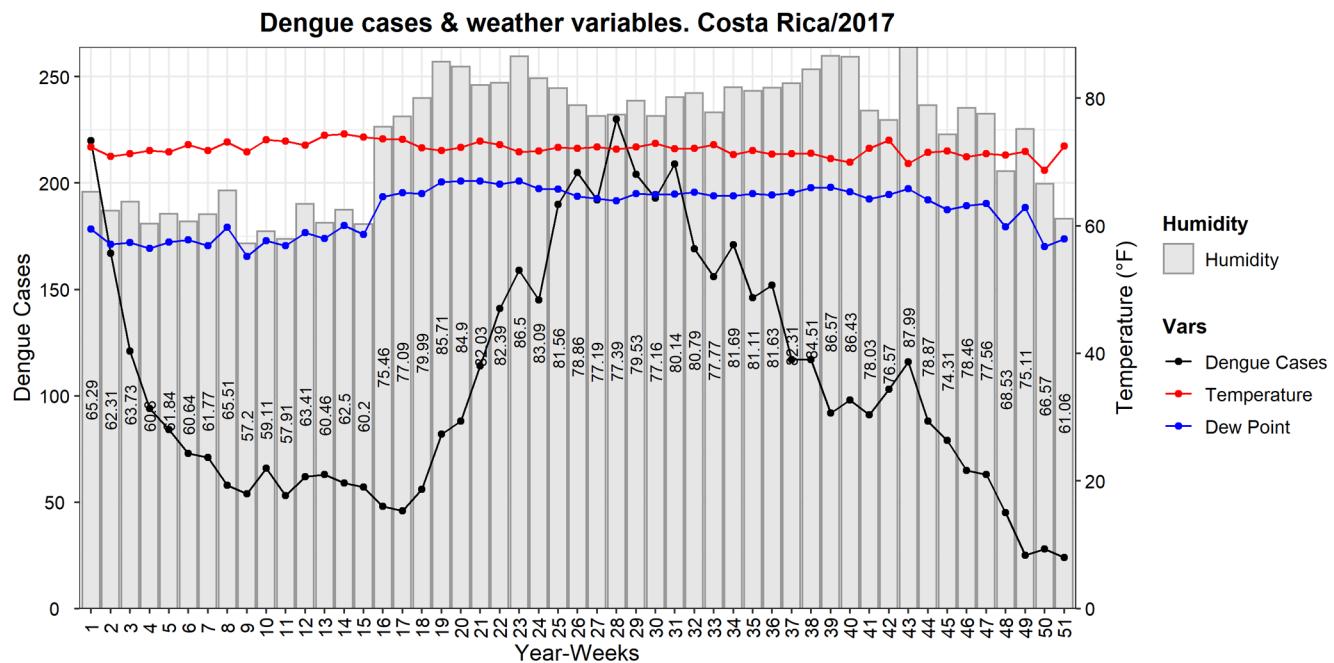


Figure 2b. Time series correlating official weekly reports of dengue cases in Costa Rica with weather variables, 2017.

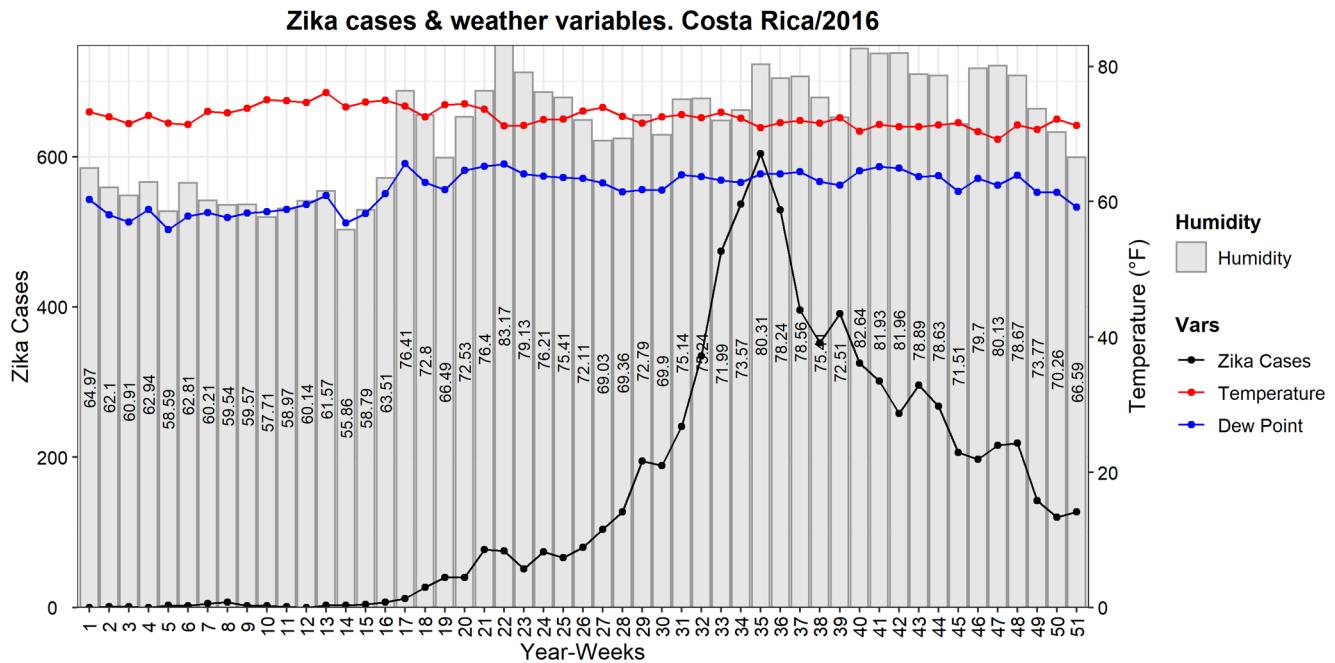


Figure 2c. Time series of Zika and weather variables for Costa Rica, 2016.

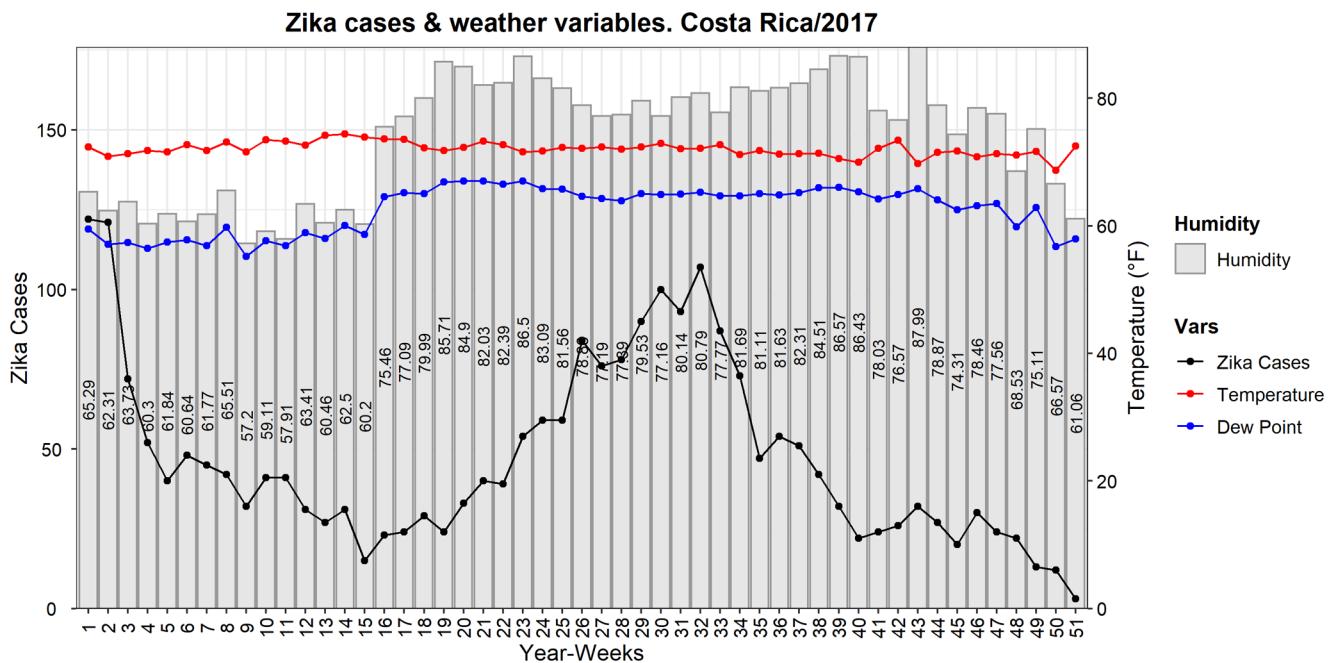


Figure 2d. Time series correlating official weekly reports of Zika cases in Costa Rica with weather variables, 2017.

graphs generated through VectorAnalytica's platform show maximum correlation between disease incidence and relative humidity and dewpoint, especially the latter. In contradiction to the literature, on the other hand, the visualized output of integrating weather variables as shown in Figures 2a-d shows that peak spikes for dengue incidence do not correspond directly with temperature spikes. But in a

remarkable and novel result that could not have been predicted or derived otherwise, analytic output clearly shows that for two years in a row, and for both Zika and dengue, there is a systematic and consistent lag between temperature and case incidence: temperature peaks around weeks 10–12, while dengue cases do not peak until approximately 20 weeks later.

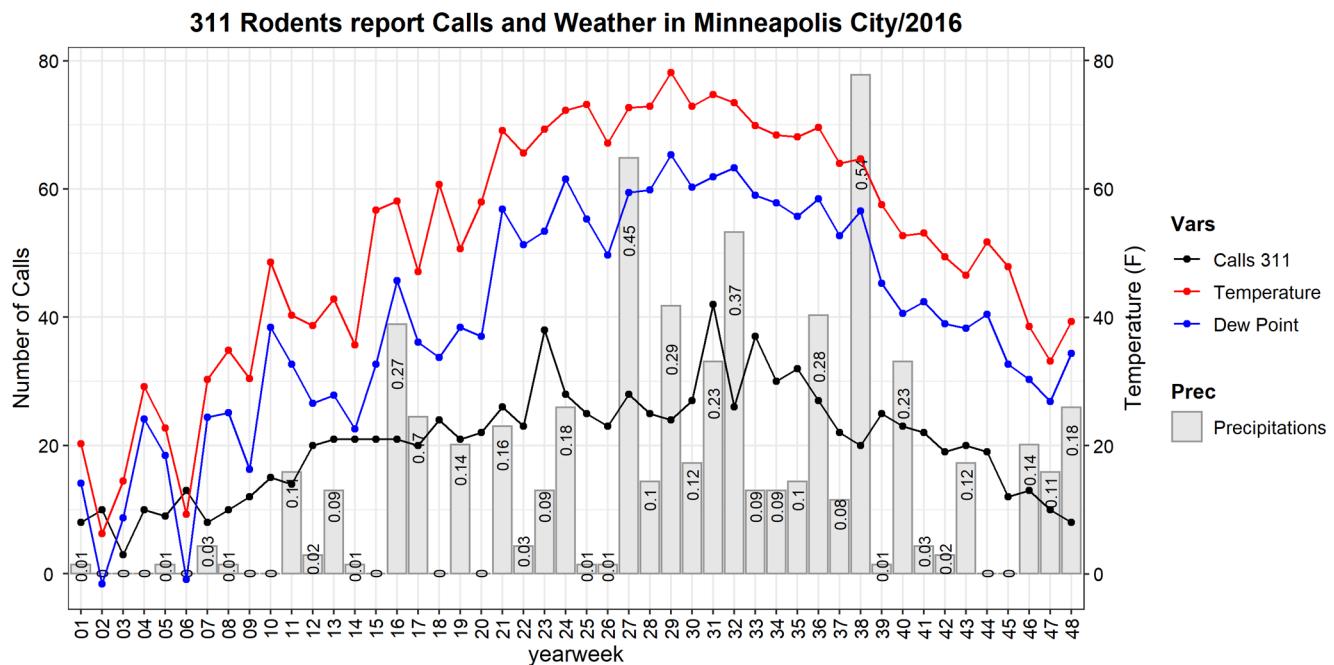


Figure 3. Time series graph of relations between rodent-related complaints to the Minneapolis 311 hotline (black line) and measures of precipitation, dew point, and temperature in the Minneapolis metropolitan area (see legend).

Case Study 3. Complaints to Minneapolis 311 hotline about rodent presence

For our final use case, we present a study of rodent infestation in Minneapolis, Minnesota. Visualization of statistical processing showed a dramatic correlation between trendlines of rodent complaints to the 311 hotline and those of dewpoint and temperature (not so much with relative humidity, possibly because of the much wider range of ambient temperatures in Minneapolis compared to Costa Rica). Remarkably and in contrast to the mosquito-borne disease cases in Costa Rica, in the case of rodent-related complaints in Minneapolis there is no time lag at all between the number of complaints and either temperature and dewpoint. In these particular conditions, for this specific vector, dewpoint and temperature can therefore serve as predictive variables; they faithfully anticipate increases in rodent complaints (as well as the presumed corresponding increase in transmission risk of rodent-related zoonotic disease).

Conclusions

The three case studies clearly show that integrated information supports decision-making and is actionable. In the

case of rodent presence in Minneapolis, private firms with this insight obtained from integrating information about dewpoint and temperature with frequency of pest complaints to 311 can systematically anticipate trends in demand for their services throughout the city. In the case of Costa Rica, it suggests the need for further study of the relation between dewpoint and humidity and dengue and Zika incidence. Such local, timely, and nuanced insight can provide public health agencies and private pest control service firms alike with data-based anticipatory awareness with which private firms can time their marketing and sales campaigns, and with which public health agencies can "anticipate to mitigate" and contain zoonotic disease spread.

An IT-based public health monitoring and tracking system today is the equivalent of a nation's highway system or the National Hurricane Center. It is the channel through which information and action can be managed to advance public health nationwide; it is a common good that must be publicly supported and disseminated.

Acknowledgments – This research was funded by Vector-Analytica, Inc.

Literature Cited

- Brooks, D.R.; Hoberg, E.P.; Boeger, W.A. 2019. The Stockholm Paradigm: Climate Change and Emerging Disease. University of Chicago Press, Chicago.
- Brooks, D.R.; Hoberg, E.P.; Boeger, W.A.; Gardner, S.L.; Galbreath, K.E.; Herczeg, D.; et al. 2014. Finding them before they find us: informatics, parasites, and environments in accelerating climate change. Comparative Parasitology 81: 155–164. <https://doi.org/10.1654/4724b.1>
- Goto, K.; Kumarendran, B.; Mettananda, S.; Gunasekara, D.; Fujii, Y.; Kaneko, S. 2013. Analysis of effects of meteorological factors on dengue incidence in Sri Lanka using time series data. PLOS One 8: e63717. <https://doi.org/10.1371/journal.pone.0063717>
- Hoberg, E.P.; Boeger, W.A.; Molnár, O.; Földvári, G.; Gardner, S.L.; Juarrero, A.; et al. 2022. The DAMA protocol, an introduction: finding pathogens before they find us. MANTER: Journal of Parasite Diversity 21. <https://doi.org/10.32873/unl.dc.manter21>
- Juarrero, A. 1999. Dynamics in Action: Intentional Behavior as a Complex System. MIT Press, Boston.
- Juarrero, A. Forthcoming 2023. Context Changes Everything: How Constraints Create Coherence. MIT Press, Boston.
- Nygren, D.; Stoyanov, C.; Lewold, C.; Måansson, F.; Miller, J.; Kamanga, A.; Schiff, C.J. 2014. Remotely-sensed, nocturnal, dew point correlates with malaria transmission in Southern Province, Zambia: a time-series study. Malaria Journal 13: 231. <https://doi.org/10.1186/1475-2875-13-231>
- Polwiang, S. 2020. The time series seasonal patterns of dengue fever and associated weather variables in Bangkok (2003–2017). BMC Infectious Diseases 20: 208. <https://doi.org/10.1186/s12879-020-4902-6>